Human Facial Emotion Classification: A Method Validation

Pedro Bessa
Faculdade de Engenharia Elétrica
Universidade Federal de Uberlândia
Uberlândia, Brazil
ORCID: 0000-0001-6412-4668

Amanda Jorge
Faculdade de Engenharia Elétrica
Universidade Federal de Uberlândia
Uberlândia, Brazil
ORCID: 0000-0001-9319-3371

Alcimar Soares
Faculdade de Engenharia Elétrica
Universidade Federal de Uberlândia
Uberlândia, Brazil
ORCID: 0000-0003-1100-3533

***Abstract*—The recognition of emotions through the perception of facial expressions can have several study applications, with applications in sociology, anthropology and, especially, in clinical and educational areas. The interpretation of human emotions is a new horizon of understanding is opened by understanding the ways people communicate and the real meaning of non-verbal behavior. The use, in literature, of controlled images considering human emotions datasets makes the proposed approaches effective indeed, but very restricted to this specific domain and controlled environments. This article aims the emotion evaluation and classification for more realistic means such as self-acquisition of images with a webcam from personal computer. And, in sum, validate a method classification, developed on previous studies by this research team, from facial coordinates. The best architecture showed an average accuracy of 46.8% among all classes. The best and worst emotions classified are in accordance with the literature.**

Keywords — Emotion Recognition, Facial Region, Machine Learning

I. Introduction

Emotions are the center of human experience, from these decisions are made and reasoning is made. It is scientific knowledge that part of the emotions in degree and intensity are represented voluntarily and, predominantly, involuntary in the non-verbal form of communication, especially through facial expressions [1,2]. The identification and rationalization of this type of communication is essential for human coexistence in society, this process is so primordial that it begins almost together with human life, being present in babies [3].

The recognition of emotions through the perception of facial expressions can have several study applications, with applications in sociology, anthropology and, especially, in clinical and educational areas, which encompass interpersonal relationships between professionals in their respective areas and patients/students [4]. A new horizon of understanding is opened by understanding the ways people communicate and the real meaning of non-verbal behavior. With the technological advance in recent years, there has been the emergence of different techniques that aim to achieve this goal through image processing.

One of the approaches is the recognition of different facial points and their distances for possible interpretations of non-verbal language. This method is of increasing relevance for the analysis and study of children on the autistic spectrum in order to compare and understand different degrees of perception in these children, in addition to providing tools for emotional identification [5]. Such techniques can also be used to support learning in virtual environments in order to understand how students, children and adults react to the conduct of education professionals and the content delivered by them. [6] In addition to being possible useful tools in the education and training of health professionals, when seeking to identify the ability of these future professionals to identify possible scenarios of sadness, pain, discomfort etc. [7].

Machine learning and computer vision hope to bring to computers the human capabilities of sensing data, understanding data, and taking action based on past and present results. Machine learning and computer vision research is still evolving. Computer vision is an essential part of the Internet of Things, Internet of Things and human brain interfaces. Complex human activities are recognized and monitored in multimedia streams using machine learning and computer vision [8].

Transfer of learning leverages the transfer of shared knowledge between training data and test data, has been proposed as a promising machine learning methodology to leverage training instances relatively less for expected prediction. Recently, designing algorithms to improve transfer learning has attracted a lot of research attention. In the literature, there are several typical methods to achieve transfer learning performance improvement [9].

As studies in the literature approach database images with an extremely controlled image acquisition environment. On the other hand, the present study focuses on the evaluation of self-taken photos by a webcam camera simulating a real environment of use and evaluation of human emotion in front of the home-machine interface. In addition, we intend to use these images of facial emotions to validate a method classification, developed on previous studies by this research team, from facial coordinates as input data for the classifiers.

II. METHODOLOGY

A. Database

The database used in validation method were facial photos self-made from the individuals of the research team. The frames were acquisitioned in static position for 5 timestamps in 30 fps making the following expressions each time: angry, disgusted, happy, neutral, sad, surprised, and afraid. In total, the frames amount were around 900 images. Each category of emotion was mapped into numerical scale. The images were self-acquisitioned, and the information was then extracted directly from the acquired images from the webcam of the computer, as illustrated on Figure 1. The images are an example of expressions happy, neutral and sad in Figure 1, A), B) and C).

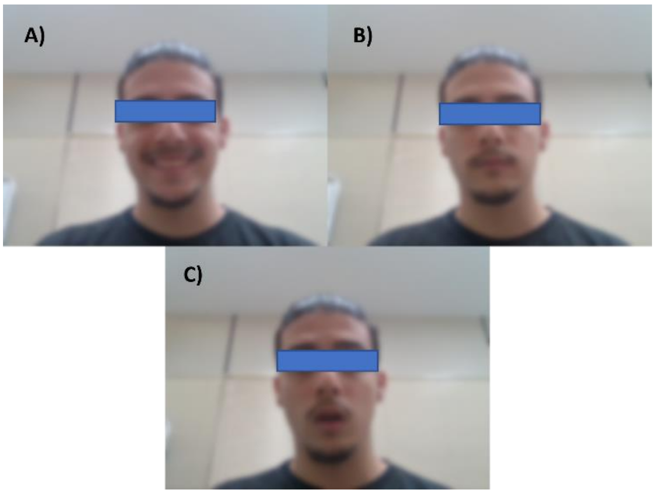

Fig. 1. Self-data acquisition, producing different expressions. Examples of happy, neutral and surprised from A), B) and C), respectively.

B. Data Preparing

The frames once saved passed through an algorithm which the information of mean pixels values, variance of pixel values from the facial region, and 7 Euclidean distances calculated from the main key points considered as presented in Figure 2. The key points are left eye (OE), right eye (OD), nose (N), left corner of the mouth (BE) and right corner of the mouth (BD). Thus, the distances considered were Right Eye-Nose (N-OD), Right Eye-Right Mouth Corner (OD-BD), Left Eye-Nose (N-OE), Left Eye, Left Mouth Corner (OE-BE), Nose-Right Mouth Corner (NBD), Nose-Left Mouth Corner (N-BE) and Right Mouth Corner-Left Mouth Corner (BD-BE).

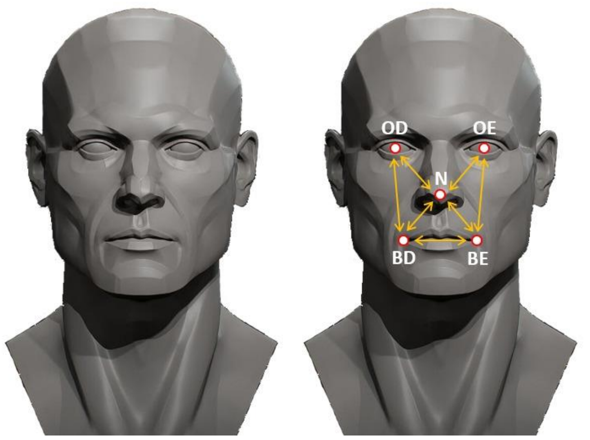

Fig. 2. Illustration of considered facial keypoints and 7 calculated Euclidean distances.

C. Data Analysis and Method Validation

On previous studies there were two types of architectures of Machine Learning: Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) to train the input data – facial coordinates with the facial image emotions. In this work, the weights from those trained models were carried out and used in this validation from webcam self-made images from the same emotions in the trained previously dataset. Thus, the acquired images were used only as new images in validation models.

III. RESULTS

After collecting and implementing the SVM and MLP analysis models and processing the data for each of the seven facial expressions, two confusion matrices were established. In both matrices the axes range from zero (0) to six (6) which are equivalent respectively to: 0 – Angry; 1 – Disgusted; 2 – Happy; 3 - Neutral; 4 – Sad; 5 – Surprised; 6 – Afraid. The Y axis (Ground Truth) corresponds in both confusion matrices to the factually collected expression and the X axis (Prediction) corresponds to the expression abstracted by means of the analysis and processing methods of the same set of photos.

In the confusion matrix from the SVM architecture, the highest hit rates can be observed in the expression related to neutral emotion (100%) and in the expression related to the emotion of surprise (54%). It is clear that, when using this method, the accuracy of the other expressions proved to be insignificant, a point to be highlighted is that there was considerable confusion of the other expressions analyzed with (3) which corresponds to the emotion Happy. The set as a whole had an accuracy of 25.6%, as shown in Figure 3.

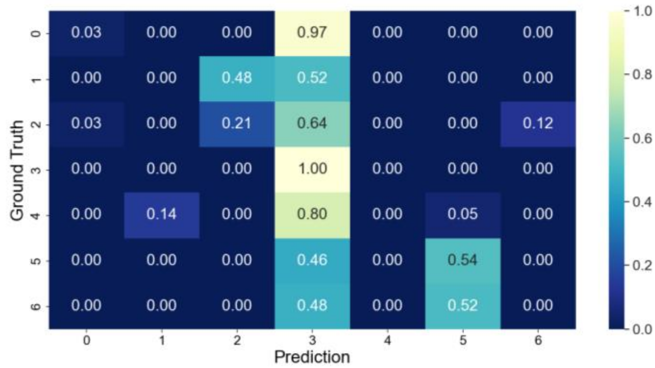

Fig. 3. Confusion Matrix from SVM architecture. Mean average accuracy 25.6%.

In the confusion matrix from the MLP architecture, the highest hit rates can be observed in the expressions related to emotions (2) Happy, (3) Neutral and (6) Afraid; with respective values of 100%, 100% and 72%. The worst results found were in the expression (0) Angry, (1) Disgusted and (4) Sad, which none obtained a prediction equivalent to factual truth. In addition to the recurrence, to a lesser extent, of the confusion of emotions (0), (1), (4) and (5) with the neutral emotion. The set had an accuracy of 46.8%, as shown in Figure 4.

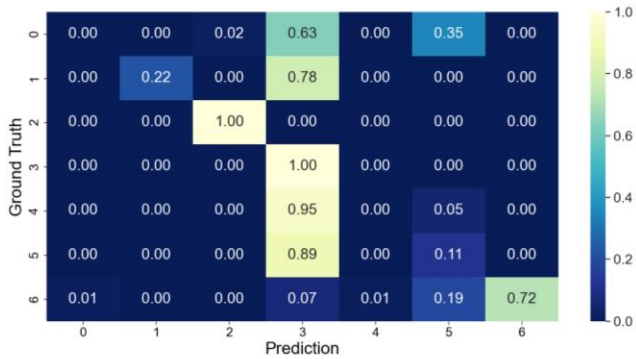

Fig. 4. Confusion Matrix from MLP architecture. Mean average accuracy 46.8%.

IV. DISCUSSION

In accordance with the state-of-art, it is showed that the best ratings for happy and neutral emotions just over 90% of the classifying algorithms [10]. The failed emotion classifications are afraid and sad emotions [10]. Our validation results showed that happy, neutral and afraid were the best accurate emotions. In this way, we can observe an upgrade in afraid emotion classification comparing with the literature. On the other hand, as presented in recent studies, the emotion sad is one of the emotions, which still has the worst accuracy among the other classes.

The use of controlled datasets or a very limited subset of these images makes the proposed approaches effective indeed, but very restricted to this specific domain. When applied to real-world problems, the overall accuracy tends to decrease significantly. This aspect can be observed mainly for the classical approaches when compared to the NNB, where a marginally better precision can be observed, justified mainly by the nature of the domain used. In general, what can be observed for the classical approaches is a very high specificity and low generality to solve the problem of emotion recognition [11].

The present study tends to move away from the sets of images collected in controlled environments and starts to focus on the evaluation for more realistic means such as acquisition of images performing different emotions. Through a webcam, the acquisition of images by the individual tends to simulate a more natural and less deterministic evaluation between man-machine interface.

Considering the vast variability of facial anatomy among different people and the complexity when using face coordinates, a possible improvement, in future studies, to be implemented is the use of an individual's neutral state of emotion as a base factor for the determination of other emotions, in order to better identify the different facial expressions [12]. Another point to be raised is the fact that emotions are exposed by facial expressions through various muscles [13]. Thus, increasing the number of points considered in the facial region to interpret anatomically and physiologically with greater precision and accuracy this type of non-verbal communication.

It is also important to mention the limitations of the presented study. Some limitations are first the need of full front region face must be detected in the image of the find the coordinates. In addition, more sophisticated emotions were not considered, for example contempt emotion, for this approach, the framework training level increases and fine facial muscles must be considered. It is recommended to use cameras with video resolution of at least 1280x720 pixels, the collections should be ideally performed with the subject near the camera (~30cm) and in an environment with good lighting and with a background not visually polluted.

V. CONCLUSION

The present study tends to move away from the sets of images collected in controlled environments and starts to focus on the evaluation for more realistic environments and image acquisitions. It is concluded that despite the use of a single individual for the collection of samples, satisfactory and positive results were obtained in resonance with the literature for neutral, happy and surprise emotions. And an improvement was also found in the rating of the afraid emotion.

Thus, the method of using facial coordinates and their respective Euclidean distances added to the application as inputs in an MLP architecture presented almost an average of 50% accuracy considering all classes. The happy, neutral and afraid classes with 100%, 100% and 72% respectively.

The results, since they were very promising, and for future work it is intended to upgrade the number of analyzed individuals and implementing a real-time application of a human emotions classifier.

ACKNOWLEDGMENT

The authors are thankful to the Federal University of Uberlândia, to the PET Biomedical Engineering and to the funding agencies Fapemig and CNPq.

REFERENCES

[1] SILVA, Maria João Martins Rosa da. A inteligência emocional como factor determinante nas relações interpessoais: Emoções, Expressões Corporais e Tomadas de Decisão. 2010. Tese de Doutorado.

[2] PORIA, Soujanya et al. Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, v. 7, p. 100943-100953, 2019.

[3] MENDES, Deise Maria Leal Fernandes; DE MOURA, Maria Lucia Seidl. Expressões faciais de emoção em bebês: importância e evidências. Estudos e pesquisas em Psicologia, v. 9, n. 2, p. 307-327, 2009.

[4] MESKA, Mateus Henrique Gonçalves et al. Reconhecimento de emoções para ambiente clínico simulado com uso de odores desagradáveis: estudo quase experimental. Revista Latino-Americana de Enfermagem, v. 28, 2020.

[5] NUNES, Gustavo Maciel; AGUIAR, Yuska Paola Costa. Emotismo: Um aplicativo para auxiliar crianças no espectro autista a reconhecer e reproduzir emoções. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. SBC, 2020. p. 692-701.

[6] DINIZ, Fábio A. RedFace-Um sistema de reconhecimento de expressões faciais para apoiar um ambiente virtual de aprendizagem. 2013. Tese de Doutorado. Dissertação de Mestrado, Universidade do Estado do Rio Grande do Norte, Mossoró, RN, Brasil.

[7] MAGON, Amanda et al. Empatia e reconhecimento de expressões faciais de emoções básicas e complexas, em estudantes de Medicina da Universidade Estadual de Mato Grosso do Sul. Research, Society and Development, v. 9, n. 9, p. e643997835-e643997835, 2020.

[8] A. I. Khan & S. Al-Habsi, Machine Learning in Computer Vision. Procedia Computer Science, 167 (2020) 1444–1451. https://doi.org/10.1016/j.procs.2020.03.355.

[9] H. Xu, W. Li, & Z. Cai, Analysis on methods to effectively improve transfer learning performance. Theoretical Computer Science, (2022). https://doi.org/10.1016/j.tcs.2022.09.023.

[10] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, & F. Ave, The extended Cohn-Kanade (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 4 (2010) 94–101.

[11] F. Z. Canal, T. R. Müller, J. C. Matias, G. G. Scotton, A. R. de Sa Junior, E. Pozzebon, & A. C. Sobieranski, A survey on facial emotion recognition techniques: A state-of-the-art literature review. Information Sciences, 582 (2022) 593–617.

[12] FILKOWSKI, Megan M.; HAAS, Brian W. Rethinking the use of neutral faces as a baseline in fMRI neuroimaging studies of axis-I psychiatric disorders. Journal of Neuroimaging, v. 27, n. 3, p. 281-291, 2017.

[13] KELTNER, Dacher et al. Emotional expression: Advances in basic emotion theory. Journal of nonverbal behavior, v. 43, n. 2, p. 133-160, 2019.